SOME ANCILLARY STATISTICS

AND THEIR PROPERTIES*

by

Robert J. Buehler

Technical Report No. 387

University of Minnesota

March, 1981

*This is a revision of Technical Report No. 364 (December 1979).

ABSTRACT

Two lists are given. The first is a list of properties of ancillary statistics; the second is a list of examples of ancillary statistics. It is then indicated which of the properties are satisfied by each example. Many of the models have the property that the parameter θ is a location parameter for the MLE $\hat{\theta}$ in every conditional distribution determined by a fixed value of the ancillary statistic. In certain other models the same state of affairs is achieved by parameter transformation. In either of these cases we call the ancillary an "exact precision index." There exist irregular models in which the precision of estimation depends not only on the ancillary but on θ as well.

1. Introduction

The purpose of this paper is to study ancillary statistics through examples. In each case a probability law $f(x;\theta)$ is assumed where x is typically a vector but θ is restricted to be a scalar.

The term ancillary statistic was introduced by Fisher (1925), and later writers unfortunately have been unable to agree completely on a definition. For purposes of the present paper it will be most convenient to adopt Basu's (1959) definition: A statistic $u(x)$ is ancillary if its distribution is the same for all θ. Basu's definition is different for example from Kendall and Buckland's (1957), which says (paraphrasing slightly): An ancillary statistic is a function of the observations that is combined with the maximum likelihood estimator to reduce the loss of information (its distribution is not mentioned). Cox and Hinkley (1974), p. 32, require in addition to Basu's condition that $(t(x),u(x))$ be minimal sufficient for some $t(x)$. A case could be made for requiring also that $t(x)$ have the same dimension as θ (dimension one in our models), or even that $t(x) = \hat{\theta}$, the maximum likelihood estimator.

According to conventional wisdom: There are difficulties with existence and uniqueness of ancillary statistics; the principle of conditionality requires us to make inferences conditional on an ancillary statistic when one exists; and an ancillary statistic by itself carries no information about θ but when used together with the maximum likelihood estimator $\hat{\theta}$, the ancillary tells us the precision of $\hat{\theta}$. In the present paper we study this conventional wisdom through examples.

The conditionality principle as stated for example by Cox and Hinkley (1974), p. 38, says (paraphrasing again): When there is an ancillary statistic, the conclusion about the parameter of interest is to be drawn as if the ancillary statistic were fixed at its observed value. In the present paper we restrict our attention to confidence intervals for a single parameter and in particular to what might be called natural confidence intervals. These are solutions which arise from the distribution of the maximum likelihood estimator in models satisfying the regularity conditions in Section 2.

The general plan of the paper is first to give a list of properties of ancillary statistics followed by a list of examples of ancillary statistics. We then indicate which examples satisfy which properties. In this way the examples are classified into main categories. In particular we introduce the concept of an "exact precision index" and show that many ancillaries have this property. Some implications for statistical inference are discussed in the conclusions.

2. Definitions

The assumed probability law will be represented by a density function $f(x;\theta)$ where x may be a vector but θ is a scalar with range $\Omega = \{\theta | \theta_L < \theta < \theta_U\}$.

D1: We will say the model $\{f(x;\theta), \theta \in \Omega\}$ (or more briefly $\{f(x;\theta)\}$) is B-regular if a unique maximum likelihood estimate $\hat{\theta}(x)$ exists for each x and if the distribution of $\hat{\theta}$ satisfies Lindley's (1958) "Condition B": The CDF $F(\hat{\theta}|\theta)$ has a derivative $\partial F/\partial\theta$ which is always negative, and $\lim F(\hat{\theta}|\theta) = 0$ (or 1) as θ tends to θ_U (of θ_L).

Except for examples EM2 and EM4 in Section 8, the present paper considers only B-regular models.

D2: If $u(x)$ is any conditioning statistic (typically an ancillary) we will say that the pair $\{f(\cdot;\cdot),u(\cdot)\}$ is B-regular if the CDF's $F(\hat{\theta}|u;\theta)$ satisfy Lindley's Condition B for all values of u.

D3: An ancillary u will be called an exact precision index (for θ) if the model $\{f(\cdot;\cdot),u(\cdot)\}$ is B-regular and satisfies property P2T (Lindley's (1958) "Condition A," transformed translation invariance) defined in Section 4 below. One purpose of the present paper is to give some justification for this term.

We will consider distributions on the parameter space some of which are fiducial distributions. To avoid getting bogged down in fiducial theory we will use the neutral term "induced distribution."

D4: For an unconditional B-regular model $\{f(\cdot;\cdot)\}$ the induced distribution of θ has density $g(\theta|x) = -\partial F(\hat{\theta}|\theta)/\partial\theta$. For a conditional

B-regular model $\{f(\cdot;\cdot),u(\cdot)\}$ the induced distribution of θ has density $g_u(\theta|x) = -\partial F(\hat{\theta}|u;\theta)/\partial\theta$. Equivalently the induced CDF's are $G(\theta|x) = 1 - F(\hat{\theta}|\theta)$ and $G_u(\theta|x) = 1 - F(\hat{\theta}|u;\theta)$.

D5: The γ percentiles of the induced distributions of D4 will be denoted by $\bar{\theta}_\gamma(\hat{\theta})$ and $\bar{\theta}_\gamma(\hat{\theta},u)$. That is, $F(\hat{\theta}|\bar{\theta}_\gamma(\hat{\theta})) = 1 - \gamma$ and $F(\hat{\theta}|u;\bar{\theta}_\gamma(\hat{\theta},u)) = 1 - \gamma$.

Thus $\bar{\theta}_\gamma(\hat{\theta})$ and $\bar{\theta}_\gamma(\hat{\theta},u)$ are unconditional and conditional upper confidence limits for θ with confidence coefficient γ.

3. Relationships between conditional and unconditional confidence limits

To be definite, let $\gamma = 0.9$ and consider the contour $\Gamma_{0.9}$ in the $(\theta,\hat{\theta})$ plane on which $F(\hat{\theta}|\theta) = 0.1$. For given $\hat{\theta}$, the θ value on this contour is $\bar{\theta}_{0.9}(\hat{\theta})$, and this is a 90 percent upper confidence limit for θ. Next fix a particular value u of a conditioning statistic $u(x)$, and consider the contour $\Gamma_{0.9,u}$ on which $F(\hat{\theta}|u,\theta) = 0.1$. We are interested in the relative orientation of Γ and Γ_u. For example, Γ_u may lie entirely on one side of Γ, or the contours may cross. Suppose for example Γ_u lies to the right and below Γ. Then $\bar{\theta}_{0.9}(\hat{\theta},u) > \bar{\theta}_{0.9}(\hat{\theta})$, and for this u, the conditional confidence interval contains the unconditional interval for all $\hat{\theta}$. Our conditional confidence in the unconditional interval is then less than 0.9. Holding u fixed we may adjust the conditional confidence level downward, plotting for example the contour $\Gamma_{0.8,u}$ on which $F(\hat{\theta}|u,\theta) = 0.2$. If this contour also lies below and to the right of $\Gamma_{0.9}$, then $\bar{\theta}_{0.9}(\hat{\theta}) < \bar{\theta}_{0.8}(\hat{\theta},u)$ and

$$P\{\theta < \bar{\theta}_{0.9}(\hat{\theta})|u,\theta\} < P\{\theta < \bar{\theta}_{0.8}(\hat{\theta},u)|u,\theta\} = 0.8$$

for all θ. This shows that the subset $\{x|u(x) = u\}$ is a negatively biased relevant subset in the sense of Buehler (1959) relative to the unconditional solution. If we can choose γ to bring the contours into coincidence, as expressed by $\bar{\theta}_\gamma(\hat{\theta},u) = \bar{\theta}_{0.9}(\hat{\theta})$, then γ is the exact conditional coverage of the unconditional solution.

4. Properties of Ancillary Statistics

In this section we list properties which may be satisfied by a B-regular model $\{f(\cdot;\cdot),u(\cdot)\}$ where $u(x)$ is ancillary.

P1: $(\hat{\theta},u)$ is minimal sufficient.

P2: Translation invariance. $(\theta_L,\theta_U) = (-\infty,\infty)$, and $F(\hat{\theta}|u;\theta) = F_o(\hat{\theta} - \theta|u)$ for some $F_o(\cdot|\cdot)$.

P3: $\mathrm{Var}(\hat{\theta}|u;\theta)$ depends on u but not on θ.

P4: For any fixed γ and u, the sign of $\bar{\theta}_\gamma(\hat{\theta},u) - \bar{\theta}_\gamma(\hat{\theta})$ is the same for all θ.

P5: For any fixed γ, $P\{\theta \leq \bar{\theta}_\gamma(\hat{\theta})|u;\theta\}$ depends on u but not on θ.

P6: The induced density $g_u(\theta|x)$ equals the posterior density for a uniform improper prior $\pi(\theta) = 1$ for $-\infty < \theta < \infty$.

PkT (k = 2,3,4,5): There exists a one-to-one increasing transformation $\tau = \Psi(\theta)$ (the same for all u) such that Pk holds with τ substituted for θ.

P6T: There exists an improper prior $\pi(\theta)$ such that the induced density $g_u(\theta|x)$ equals the posterior density.

5. Discussion of Properties

P1 is a strengthening of Cox and Hinkley's requirement mentioned in Section 1 which restricts $t(x)$ to be the MLE. Possibly P1 was actually intended by Fisher. For our purposes it would do about as well to replace "minimal sufficient" by "sufficient" for the following reason: If $u = (u_1,u_2)$ is ancillary and $(\hat{\theta},u_1)$ is minimal sufficient, then under an assumption of completeness, u_2 is independent of $(\hat{\theta},u_1)$, so that the additional conditioning on u_2 is irrelevant. This is the case in several of our examples.

P2T states that the distribution $F(\hat{\theta}|u;\theta)$ satisfies Lindley's (1958) "Condition A" for all u with the transformation $\Psi(\theta)$ the same for all u.

P3 is an attempt to formalize the statement that the ancillary statistic determines the precision of θ.

P4 expresses the non-crossing of contours Γ and Γ_u as described in Section 3.

P5 states that the conditional confidence in the unconditional confidence intervals is constant and thus the ancillary defines relevant subsets as mentioned in Section 3.

6. Relationships and Implications of the Properties

Proposition 1. P2T implies that the contours in the $\theta,\hat{\theta}$ plane on which $F(\hat{\theta}|u;\theta) = \gamma$ are a one parameter family which never cross each other for any u or any γ. The contours of $F(\hat{\theta}|\theta) = \gamma$ belong to the same one parameter family.

Proof. By P2T there exists $\tau = \Psi(\theta)$ such that $F(\hat{\tau}|u;\tau) = F_o(\hat{\tau} - \tau|u)$. The contours in the $\tau,\hat{\tau}$ plane are $\hat{\tau} = \tau + c$, $-\infty < c < \infty$. Averaging over the distribution of u (using that u is ancillary), $F(\hat{\tau}|\tau) = F_o(\hat{\tau} - \tau)$ so that the family of contours is the same. Transforming to $\theta = \Psi^{-1}(\tau)$ and using $\hat{\theta} = \Psi^{-1}(\hat{\tau})$, we get a one-parameter family of noncrossing contours $\hat{\theta} = \Psi^{-1}(\Psi(\theta) + c)$, $-\infty < c < \infty$.

Proposition 2. Let $v = \partial \log f(x;\theta)/\partial\theta$ be the score function and let $I(\theta) = \text{Var } v$ and $I(\theta,u) = \text{Var } (v|u)$ be the unconditional and conditional Fisher information. Then: (a) P1 and P2 imply $I(\theta)$ and $I(\theta,u)$ are free of θ, and (b) P1 and P2T imply that $I(\theta,u) = I_1(\theta)I_2(u)$ for some functions I_1,I_2.

Proof. (a) Evident from sufficiency and invariance. (b) By P2T and (a), $I(\tau,u) = I_o(u)$. Transforming by $\theta = \Psi^{-1}(\tau)$, $I(\theta,u) = I_o(u)(d\Psi/d\theta)^2$.

Proposition 2 shows that when $(\hat{\theta},u)$ is a sufficient statistic, a necessary condition for the ancillary u to be an exact index of precision (D3, Sec. 2) is that the conditional Fisher information factor into a function of θ times a function of u.

Proposition 3. The following implications hold:

(i) $Pk \Rightarrow PkT$ for $k = 2, 3, 4, 5, 6$.

(ii) $P4T \Rightarrow P4$, $P5T \Rightarrow P5$.

(iii) $P2 \Rightarrow Pk$ and $P2T \Rightarrow PkT$ for $k = 3, 4, 5$.

(iv) If P1 then $P2 \Leftrightarrow P6$ and $P2T \Leftrightarrow P6T$.

Proof.

(i) This is trivial.

(ii) The proof is similar to that of Proposition 1 where we note that contours in the $\theta,\hat{\theta}$ plane are transformed into contours in the $\tau,\hat{\tau}$ plane by $\tau = \Psi(\theta)$, $\hat{\tau} = \Psi(\hat{\theta})$, leaving the relative orientation unchanged. For the second part, $\tau \leq \bar{\tau}_\gamma(\hat{\tau})$ and $\theta < \bar{\theta}_\gamma(\hat{\theta})$ are the same event.

(iii) $P2 \Rightarrow P3$ is evident. As in the proof of Proposition 1, the contours of $F(\hat{\theta}|\theta)$ and $F(\hat{\theta}|u,\theta)$ are all 45 degree lines in the $(\theta,\hat{\theta})$ plane. Both P4 and P5 follow from this. If P2T holds, then the transformation involved in P2T is the one which establishes P3T, P4T, P5T.

(iv) The proofs here are essentially those of Lindley (1958). Assume P2. Then $\partial F(\hat{\theta}|u,\theta)/\partial\hat{\theta} = -\partial F(\hat{\theta}|u,\theta)/\partial\theta$. The former expression is the density $f(\hat{\theta}|u,\theta)$ while the latter is by definition the induced density $g_u(\theta|x)$. Using P1 and ancillarity of u, the likelihood given x is proportional to $f(\hat{\theta}|u;\theta)$. For a uniform prior the posterior is proportional to the same function, and the normalization constant is unity by Condition P2. This establishes P6. Next assume P6. As above, the posterior for a uniform prior is $f(\theta|u;\theta) = \partial F(\hat{\theta}|u;\theta)/\partial\hat{\theta}$. By P6 this equals the induced density $g_u(\theta|x) = -\partial F(\hat{\theta}|u,\theta)/\partial\theta$. This yields the differential equation $\partial F(\hat{\theta}|u,\theta)/\partial\hat{\theta} = -\partial F(\hat{\theta}|u,\theta)/\partial\theta$ whose general solution is $F(\hat{\theta}|u,\theta) = F_o(\hat{\theta} - \theta|u)$ for some function $F_o(\cdot|\cdot)$, establishing P2.

Next assume P2T. Then there is $\tau = \Psi(\theta)$ such that τ is a location parameter for $\hat{\tau}$ in each conditional distribution. By the argument above, the induced density of τ is the posterior for a uniform prior on τ. From this we get P6T by noticing that in transforming from τ to θ both the posterior and induced densities are modulated by multiplying by the same Jacobian $d\Psi/d\theta$.

Finally assume P6T. The prior $\pi(\theta)$ can be used to define a transformation $\tau(\theta)$ such that $\pi(\theta) = d\Psi/d\theta$, giving a uniform prior on τ. For this choice of parameter P6 holds, and we have shown P6 implies P2. Thus P6T implies P2T.

7. Notation for distributions

$N(\mu,\sigma^2)$ denotes normal with mean μ, variance σ^2.

$G(\alpha,p)$ denotes a gamma distribution with density $f(x;\alpha,p) = \alpha^p x^{p-1} e^{-\alpha x}/\Gamma(p)$ $(x > 0,\ \alpha > 0,\ p > 0)$.

$\mathrm{Exp}(\theta)$ denotes an exponential distribution with density $f(x,\theta) = \theta e^{-\theta x}$ $(x > 0,\ \theta > 0)$. $\mathrm{Exp}(\theta) = G(\theta,1)$.

$\mathrm{Lind}(\theta)$ denotes what we will call a Lindley (1958) distribution with density $f(x,\theta) = \theta^2(\theta+1)^{-1}(x+1)e^{-\theta x}$ $(x > 0,\ \theta > 0)$ (see Appendix E).

$\chi_2^2(\lambda)$ denotes a noncentral chi square distribution with two degrees of freedom and noncentrality λ, whose density is (see for example Graybill 1976, p. 125)

$$f(x;\lambda) = 2^{-j-1}e^{-\lambda-x/2}\sum_{j=0}^{\infty}\lambda^j x^j/(j!)^2 \qquad x > 0$$

8. Examples of Ancillary Statistics

In each model the distribution of the ancillary is given either implicitly or explicitly. It should be noted that the properties we are considering would not change with a change in the distribution of the ancillary, and making an explicit assumption about this distribution serves only to help us focus on a specific model.

The following notation is convenient. E = example, L = location parameter model, S = scale parameter model, X = exponential model, I = irregular model, M = miscellaneous model, g = generalized, n = sample of size n. Thus ES1gn denotes scale parameter example number one, generalized, with sample size n.

EL1: Two measuring instruments (Cox, 1958). $P(u = 0) = P(u = 1) = 1/2$; $\{x|u\} \sim N(\theta,\sigma_u^2)$, $\hat{\theta} = x$.

EL1n: Sample $(x_1,u_1),\ldots,(x_n,u_n)$ from EL1. $\hat{\theta} = (\Sigma\sigma_{u_j}^{-2} x_j)(\Sigma\sigma_{u_j}^{-2})^{-1}$, (Efron and Hinkley, 1978).

EL1g: In EL1 replace normal densities by two arbitrary location models $f(x|u = 0,\theta) = f_o(x - \theta)$, $f(x|u = 1,\theta) = f_1(x - \theta)$.

EL1gn: n observations from EL1g.

EL2: Fisher-Pitman location model. $f(\underset{\sim}{x};\theta) = \prod_{i=1}^{n} f(x_i - \theta)$. $u = (x_{(1)} - x_{(2)}, x_{(1)} - x_{(3)},\ldots,x_{(1)} - x_{(n)})$, the spacings of the ordered observations $x_{(i)}$.

EL2g: Location model assuming neither independence nor identical distributions. $f(x;\theta) = f(x_1 - \theta,\ldots, x_n - \theta)$. $u = (x_1 - x_2,\ldots,x_{n-1} - x_n)$, spacings of unordered observations.

EL3: One-parameter normal regression. $u \sim N(0,1),\{x|u\}\sim N(\theta u,1)$.

Observe $(u_1,x_1),\ldots,(u_n,x_n)$, $\hat\theta = \Sigma u_i x_i/\Sigma u_i^2$, $\{\hat\theta|(u_1,\ldots,u_n)\} \sim N(\theta,(\Sigma u_i^2)^{-1})$.

EL4: Sprott's (1961) ancillary. $x_1 \sim N(n\theta,n)$, $x_2 \sim G(m,ce^{k\theta})$, $u = x_1/n - (\log x_2)/k$.

ES1: Two-valued ancillary with reciprocal exponentials. $P(u = 0) = P(u = 1) = 1/2$. $\{x|0\} \sim \mathrm{Exp}(\theta)$, $\{x|1\} \sim \mathrm{Exp}(\theta^{-1})$.

ES1n: n observations from ES1.

ES1g: $f(x|u = 0,\theta) = \theta f(\theta x)$, $f(x|u = 1,\theta) = \theta^{-1}f(x/\theta)$.

ES2: Fisher-Pitman scale model. $f(x;\theta) = \theta^{-n}\prod_{i=1}^{n} f(x_i/\theta)$, $\theta > 0$, $x_i > 0$, u = quotients of ordered observations.

ES2g: $f(x,\theta) = \theta^{-n}f(x_1/\theta,\ldots,x_n/\theta)$, u = quotients of x_i's (unordered).

ES3n: Fisher's gamma hyperbola (Fisher, 1973, p. 169; Efron and Hinkley, 1978, example 3.2). $x_1 \sim \mathrm{Exp}(\theta)$, $x_2 \sim \mathrm{Exp}(\theta^{-1})$. Observe n pairs x_1, x_2. If $s_j = \sum_{i=1}^{n} x_{ji}$, $j = 1, 2$, then $\hat\theta = (1/2)\log(s_1/s_2)$ and $u = s_1 s_2$.

ES3gn: Sample of size n from $f(x_1,x_2,\theta) = f(\theta x_1,x_2/\theta)$, $\theta > 0$, where $f(z_1,z_2)$ is a density on $0 < z_1 < \infty$, $0 < z_2 < \infty$. The ancillary statistic can be represented by the n products $x_{1i}x_{2i}$ together with $n - 1$ quotients of the ordered x_i's: $x_{(1)}/x_{(2)},\ldots,x_{(n-1)}/x_{(n)}$.

ES4n: Normal with known coefficient of variation (Hinkley, 1977). $x \sim N(\theta,c^2\theta^2)(\theta > 0$, c known). $u = s_2^{1/2}/s_1$, $s_k = n^{-1}\Sigma x_j^k$, $\hat\theta = (1/2)s_1\{(1 + 4u^2)^{1/2} - 1\}$.

ES4gn: Arbitrary shape with known coefficient of variation. $x = \theta y$ where y has density $g(y)$ for $-\infty < y < \infty$. Then $f(x) = \theta^{-1}g(x/\theta)$. The ancillary statistic gives the number of negative observations and the quotients of ordered positive and negative observations separately.

EX1: An exponential model mentioned by Barndorff-Nielsen (1980) (see Appendix F). One bivariate observation x,y from density

$$f(x,y;\theta) = c\theta I_o(2\sqrt{xy})e^{-cx - (x/\theta) - \theta y} \qquad x > 0,\ y > 0,\ \theta > 0.$$

Here $u = x \sim \mathrm{Exp}(c)$.

EI1: $P(u = 0) = P(u = 1) = 1/2$. $\{x|0\} \sim N(\theta,1), \{x|1\} \sim N(\theta^3,1)$.

EI2: $P(u = 0) = P(u = 1) = 1/2$. $\{x|0\} \sim \mathrm{Exp}(\theta), \{x|1\} \sim \mathrm{Lind}(\theta)$.

EM1: Fisher's normal circle (Fisher 1973, p. 138, Efron and Hinkley 1978, p. 464). $x_1 - \rho\cos\theta$, $x_2 - \rho\sin\theta$ are independent $N(0,1)$, where ρ is known. The ancillary is $u = (x_1^2 + x_2^2)^{1/2}$.

EM2: One observation (x_1,x_2) from a bivariate normal distribution with zero means, unit variances and correlation θ. $u = x_1$.

EM2n: Sample of size n from EM2.

EM3: (Basu 1959). Two observations (x_1,x_2) from $N(\theta,1)$. $u = x_1 - x_2$ if $x_1 + x_2 < c$ and $u = x_2 - x_1$ if $x_1 + x_2 \geq c$.

EM4: (Basu 1964, Cox 1971, Barnard and Sprott 1971). Multinomial distribution with four cells whose frequencies are $x_1,\ldots,x_4$ and whose probabilities are $p_1 = (1 - \theta)/6$, $p_2 = (1 + \theta)/6$, $p_3 = (2 - \theta)/6$, $p_4 = (2 + \theta)/6$, where $-1 \leq \theta \leq 1$. $u_1 = x_1 + x_2$ and $u_2 = x_1 + x_4$ are separately but not jointly ancillary.

9. Discussion of the Examples

We give two general results before discussing the models individually.

Proposition 4: Subject to B-regularity (see D1, Section 2), all of the EL models satisfy property Pk for $k = 2, 3, 4, 5, 6$.

Proof. Verification of P2 is reasonably straightforward in each case, and the rest follow from Proposition 3.

Proposition 5: Subject to B-regularity, all of the ES models satisfy PkT for $k = 2, 3, 4, 5, 6$.

Proof. It can be shown that the transformation $\tau = \log\theta$ reduces each ES model to a location model.

9.1. Location Models

EL1 is occasionally put forward in support of the principle of conditionality--if we know which of two measuring instruments was used, our inference about θ should be conditional on this information.

EL1n has been discussed by Efron and Hinkley (1978) as an example of combining information and determination of the relevant conditional variablitiy of the combined estimator.

In EL1g the induced density (see D6, Section 2) is $g_u(\theta|x) = f_u(x - \theta)$.

In EL1gn supose $u = (u_1,\ldots,u_n)$ contains r zeros and $s = n - r$ ones. For fixed r we have r observations from location model f_o and s from model f_1. This falls within the generalized Fisher-Pitman model (Appendix A). The induced distribution is a posterior distribution for a uniform prior conditionally for each fixed r and hence also unconditionally.

EL2 and EL2g are discussed in Appendix A.

EL3 is simpler than the usual two-parameter model whose conditional properties have been discussed by Fisher (1973), pp. 86-89.

Sprott's example, EL4, falls within the general location model theory as indicated in Appendix D since θ is a location parameter for $(1/k)\log x_2$.

9.2. Scale Models

ES1 is clearly a variant of Cox's example EL1. Unconditionally x has the mixture distribution $f(x;\theta) = \frac{1}{2}(\theta e^{-\theta x} + \theta^{-1}e^{-x/\theta})$, but this distribution should not be used, even by disbelievers in the conditionally principle, as it is not the most natural procedure. One reason is that $f(x;\theta)$ is not B-regular. But more basically we want the analog of the procedure used in more subtle models where the ancillary may be hard to recognize. In such cases the natural procedure is to find first the MLE $\hat{\theta}$ and then its distribution. Thus we have $\hat{\theta} = x^{-1}$ is $u = 0$ and $\hat{\theta} = x$ if $u = 1$, and the MLE has unconditional CDF $F(\hat{\theta}|\theta) = \frac{1}{2}(1 - e^{-\hat{\theta}/\theta} + e^{-\theta/\hat{\theta}})$, which is seen to be a scale model. An alternative analysis leading to the same result, would consist in transforming first to a location model, as in the following paragraph. The situation here differs from that in Cox's example, EL1, in an interesting way. In EL1 the induced distributions for θ for $u = 0,1$ ($\hat{\theta}$ fixed) differ in variance but not in mean. In ES1 the induced distributions for $\log\theta$ for $u = 0,1$ ($\hat{\theta}$ fixed) are stochastically ordered (because f_o and f_1 defined in the next paragraph are). Thus lower and upper confidence limits shift <u>in the same direction</u> as u varies (and the unconditional limits of course take intermediate values). The conditional Fisher information (given u) does not depend on u.

To analyze ES1n, first transform by $\tau = \log\theta$, $v = -\log x$ if $u = 0$ and $v = \log x$ if $u = 1$. This reduces the problem to a location model already considered in EL1gn, with $f_o(y) = \exp(-y - e^{-y})$ and $f_1(y) = \exp(y - e^{-y})$.

Similarly ES2 and ES2g are transformed into EL2 and EL2g by taking logs.

ES3 and ES3g are discussed in Appendix B.

ES4 and ES4g are discussed in Appendix C.

9.3. The Exponential Model

This model is discussed at length in Appendix F. Since the conditional Fisher information $I_{y|x}(\theta) = \theta^{-2} + 2x\theta^{-3}$ is not a function of θ times a function of x we know from Proposition 2 (Section 6) that P2T fails and x is not an exact precision index (D3, Section 2). The parameter θ is in some sense an approximate scale parameter and $\tau = \log\theta$ is an approximate location parameter. The conditional information relative to τ is $I_{y|x}(\tau) = 1 + 2xe^{-\tau}$ and the unconditional information is $f_{x,y}(\tau) = 1 + 2ce^{-\tau}$. Presumably still "closer" to an unconditional location parameter would be $\lambda(\tau) = \int_{-\infty}^{\tau}(1 + 2ce^{-t})^{\frac{1}{2}}dt$ (or replace c by x to get an approximate conditional location parameter).

It is reasonable to call x an approximate index of precision. To support this claim we refer to numerical calculations which show that as x is increased, conditional confidence intervals become shorter (Appendix F). Qualitatively this agrees with the fact that the conditional Fisher information increases with x for any τ. Numerical calculations also show that as both x and the confidence level are varied the contours defining the conditional confidence limits do not constitute a one-parameter family but intersect.

9.4. Irregular Models

EI1 and EI2 are deliberately pathological counterexamples. In EI1 different functions of θ serve as location parameter depending on the value of u. Since no single transformation of θ gives a location parameter for both values of u, we have a violation of P2T. In EI2 P2T is also violated, but for a different reason: It is known that no transformation of θ in the Lindley distribution yields a location model. On a log-log plot the contours of $F(\hat{\theta}|u = 0,\theta)$ ($u = 0$ gives $\mathrm{Exp}(\theta)$) are parallel lines having unit slope. For $F(\hat{\theta}|u = 1,\theta)$ ($\mathrm{Lind}(\theta)$ case) the contours are curvilinear, concave upward (see Appendix E for some calculations). The crossing of the straight and curvilinear contours violates properties like P4.

Returning to EI1, we may think qualitatively as follows. Suppose $\theta = 10$. If $u = 0$ then most likely we will observe $8 < x < 12$, and a 95% confidence interval would have end points $x \pm 2$, approximately. But if $u = 1$, then we observe $998 < x < 1002$ and a typical 95% confidence interval would be $998^{1/3} < \theta < 1002^{1/3}$, much narrower than above. Thus when $\theta = 10$, $u = 1$ gives much more precision than $u = 0$. But when $\theta = 0$ the reverse is true. Thus in this case the value of u does not give an index of precision. The value of unknown θ must also be taken into account.

We can also think about EI1 from the point of view of one-sided confidence limits. Take the one-sigma value $\gamma = 0.84$. For $u = 0$ the upper confidence limit is $\hat{\theta} + 1$, while for $u = 0$ the contours $F(\hat{\theta}|0,\theta)$ have equation $\hat{\theta}^3 - \theta^3 = \text{constant}$ and the corresponding upper confidence limit is $(\hat{\theta}^3 + 1)^{1/3}$, which is larger than $\hat{\theta} + 1$ when $-1 < \hat{\theta} < 0$, and smaller otherwise. If we were to calculate an unconditional 84 percent upper confidence limit we would find that the conditional viewpoint would assess it as too large or too small depending not only on the value of u, but on the value of $\hat{\theta}$ as well ($u = 0$ and $\hat{\theta} \in (-1,0)$ or $u = 1$ and $\hat{\theta} \notin (-1,0)$, too large; otherwise too small). For this reason the ancillary statistic here is a disappointment to us in that it fails to serve as a precision index.

Similar conclusions about EI1 are available from the Fisher information. The conditional Fisher information is either 1 or $9\theta^4$ according as $u = 0$ or 1. Thus when θ is close to zero, $u = 0$ gives more information than $u = 1$, but the situation is reversed when $|\theta|$ is large.

9.5. Miscellaneous Models

The following models have all appeared in earlier literature. Our purpose is to show why each fails in some sense to conform to conditions imposed in the present paper.

EM1 fails to fall in the EL (location) category only because θ defines points on a circle rather than on $(-\infty,\infty)$. EM1 does exhibit all the desirable properties of the EL models suitably restated for the circle.

EM2 is a standard example in which x_1, x_2 are separately but not jointly ancillary. The ancillary x_1 (or x_2) is of little help for inference for two reasons: (i) $(\hat{\theta},x_1)$ is not sufficient, so that P1 is violated, and (ii) $\{f(x;\theta),x_1\}$ is not B-regular.

EM2n has been considered by Efron and Hinkley (1978), Section 6. It is not known whether there exists an ancillary u such that $(\hat{\theta},u)$ is sufficient.

EM3 is of interest in exhibiting nonuniqueness of ancillaries, but in fact it has little implication for inference. The MLE $\hat{\theta}$ alone is sufficient so that the conditional induced distribution would not differ from the unconditional one for any ancillary u (any value of c).

EM4 falls outside the primary framework of this paper because the distribution of $x_1,\ldots,x_4$ is discrete, so that we cannot obtain an induced distribution by a pivotal argument, except perhaps in some large-sample approximation. A second difficulty however is that neither $(\hat{\theta},u_1)$ nor $(\hat{\theta},u_2)$ is not sufficient for general $n = \Sigma x_i$, so that P1 is violated. (Cox (1971) points out that u_1 is a component of a minimal sufficient statistic, but does not consider the sufficiency of $(\hat{\theta},u_1)$. Compare our remarks in Section 1 on the Cox-Hinkley definition of ancillarity.) To see this take $n = 4$. Then $(x_1,\ldots,x_4) = (0,0,3,1)$ and $(0,0,4,0)$ both give $(\hat{\theta},u_1) = (-1,0)$ but have different likelihoods. For $n = 1$, $(\hat{\theta},u_1)$ is minimal sufficient, as Basu (1964) pointed out.

10. Discussion and Conclusions

The traditional role of an ancillary statistic u is to provide a reference set for inference. To sharpen the focus we have restricted inference procedures to confidence intervals for a single parameter θ. A further restriction is that the model be sufficiently regular to give an induced distribution (or "confidence distribution") - essentially a nested family of confidence procedures, one for each possible value of the confidence coefficient γ.

It is widely recognized that any statistic, u, defining a reference set for inference about θ should have a distribution which is free of θ because the conditioning would otherwise ignore information in the conditioning variable.

The present paper has focussed on the concept of an "exact precision index." For this we consider induced distributions obtainable from conditional pivots involving the MLE, $\hat{\theta}$. After restricting by suitable regularity conditions, we find that examples fall into three main categories. The first category contains translation invariant models which exhibit several properties: The shape of the distribution of $\hat{\theta}$ given (θ,u) depends on u but not on θ. The shape of the induced distribution of θ given $(\hat{\theta},u)$ depends on u but not on $\hat{\theta}$. The conditional Fisher information given u typically depends* on u but never on θ. For unconditional confidence intervals based on the distribution of $\hat{\theta}$, subsets with fixed values of u are relevant reference sets in the sense of Buehler (1959).

---

*In ES1, Section 8, the conditional Fisher information is constant over u.

The second category contains models transformable to the first by parameter transformation. For these the relevant reference set property just mentioned continues to hold, and the conditional Fisher information factors into a part depending only on θ times a part depending only on u.

For models in either of the first two categories we have called the ancillary statistic an exact precision index.

All other regular models are placed in a third category. For these we may think of the precision of estimation as dependent not only on u but on θ as well, a property which must remain true no matter how θ is transformed. One may also think of the estimated precision as depending not only on u but also on $\hat{\theta}$. For these models the conditional Fisher information will typically involve θ and u in a way which cannot be separated by factorization. Qualitatively we can think of the ideal behavior as failing to various degrees. In our example of an exponential ancillary, the ancillary in fact seems to provide an approximate measure of precision.

The question of approximate ancillarity (distribution weakly dependent on θ) was raised by Cox and Hinkley (1974), p. 34, and has been the subject of recent research (for example Efron and Hinkley (1978), Cox (1980), Hinkley (1980), Barndorff-Nielsen (1980)). We suggest that a systematic approach to the study of any exact or approximate ancillary might begin by putting the parameter in standard form: $\tau = \Psi(\theta)$ where $(d\Psi/d\theta)^2 = I(\theta)$, the unconditional Fisher information. (Hinkley (1980) has found this step to be important in relating the likelihood function to approximate ancillaries.) In reasonably regular models this is always possible in principle and does not depend on any proposed exact or approximate ancillary. If there exists an ancillary u which is an exact index of precision, then by the arguments in Section 6 and Lindley (1958), $[\partial F(\hat{\theta}|u;\theta)/\partial\hat{\theta}]/[\partial F(\hat{\theta}|u;\theta)/\partial\theta]$ must be expressible as $a(\hat{\theta})/b(\theta)$ and $\Psi(\theta)$ could alternatively be obtained from $d\Psi/d\theta = b(\theta)$. It is straightforward to show that the unconditional $F(\hat{\theta}|\theta)$ could be substituted for $F(\hat{\theta}|u;\theta)$ with the same result. Clearly the "CDF method" (using $b(\theta)$) is less general than "information method" (using $I(\theta)$) since it depends on the factorization $a(\hat{\theta})/b(\theta)$, but when applicable it would lead to the same result.

Once the parameter is in standard form, any exact or approximate ancillary can be studied to see whether the shape of the distribution $\hat{\theta} - \theta$ given u is approximately free of θ for each u. If so, it would be reasonable to call u an approximate index of precision.

Nonuniqueness of ancillary statistics has been considered a weakness of theories of conditional inference. Within the restrictive assumptions arbitrarily chosen for the present study, all nonuniqueness examples known to the writer have been ruled out. The relevant restrictions are: (1) continuous distributions, (2) sufficiency of $(\hat{\theta},u)$, (3) invertibility of distribution of $\hat{\theta}$ given u (B-regularity; see Section 2). This suggests a conjecture: There exists no model such that $(\hat{\theta},u_1)$ is sufficient, $(\hat{\theta},u_2)$ is sufficient, u_1 and u_2 are each ancillary, B-regularity holds in each case, and the induced distributions of $\hat{\theta}$ are different for the two ancillaries.

Finally some remarks on the conditionality principle as stated for example by Cox and Hinkley (1974), p. 38. The above conjecture if true would remove nonuniqueness objections when the inference consists of an induced distribution and the stated regularity conditions are imposed. And how does conditionality relate to the concept of an exact precision index? The conditionality argument seems sensible whether or not the ancillary is an exact precision index. When it is not, the loss is to the tidiness of interpretation more than to the conditionality principle. Finally we remark that it is not at all clear at this time how the conditionality principle should be amended to apply to approximate ancillaries.

Appendix A

Generalized Fisher-Pitman Fiducial Distributions

The Fisher-Pitman theory of location and scale models applies to random samples from location models, scale models, and joint location-scale models having respectively the densities $f(x - \theta)$, $\sigma^{-1}f(x/\sigma)$ and $\sigma^{-1}f((x - \theta)/\sigma)$. There is no difficulty in generalizing the main results of this theory to the case of nonidentically distributed observations (as with likelihood $\Pi f_i(x_i - \theta)$ replacing $\Pi f(x_i - \theta)$) and moreover to dependent observation, as with likelihood $f(x_1 - \theta,\ldots,x_n - \theta)$. Indeed these cases fall within the scope of the invariant models considered by Fraser (1961 a,b) and Hora and Buehler (1966). The principal results needed for our present purposes are that the fiducial distribution is a posterior distribution corresponding to prior measure equal to right Haar measure ($d\theta$, $d\sigma/\sigma$ and $d\theta d\sigma/\sigma$ in the three cases cited) and the fiducial limits are confidence limits obtained from a pivotal quantity conditional on an appropriate ancillary statistic. For the model $f(x_1 - \theta,\ldots,x_n - \theta)$, $u = (x_1 - x_2, x_1 - x_3,\ldots,x_1 - x_n)$ is an appropriate ancillary. For the model $\Pi f(x_i - \theta)$ the order statistic $(x_{(1)},\ldots,x_{(n)})$ is sufficient and it is possible to make a sufficiency reduction of u to $u^* = (x_{(1)} - x_{(2)},\ldots,x_{(1)} - x_{(n)})$, but this is not essential since we get the same induced distribution either way. Similar considerations apply to intermediate models such as

$$\text{(A.1)} \qquad \prod_{i=1}^{r} f_o(x_i - \theta) \prod_{i=r+1}^{n} f_1(x_i - \theta).$$

Appendix B

Fisher's Gamma Hyperbola and Generalizations

Fisher (1973) considers the joint density $f(x,y;\theta) = \exp(-\theta x - y/\theta)$. Efron and Hinkley (1978) call this "Fisher's Gamma Hyperbola." Under the transformation $\tau = \log\theta$, $u = -\log x$, $v = \log y$ we find

$$(B.1) \qquad g(u,v;\tau) = g_1(u-\tau)g_2(v-\tau)$$

where $g_1(y) = \exp\{-y - e^{-y}\}$ and $g_2(y) = \exp\{y - e^{y}\}$. This falls within the scope of generalized Fisher-Pitamn models (Appendix A). For one bivariate observation (u,v) the ancillary is $u - v = -\log(xy)$, or equivalently xy. For n observations the statistic $(u_1-u_2,\ldots,u_{n-1}-u_n, u_n-v_1, v_1-v_2,\ldots,v_{n-1}-v_n)$ is ancillary, but a minimal sufficient reduction brings this down to $(\Sigma x_i)(\Sigma y_i)$.

The generalization ES3g assumes $f(x,y;\theta) = f(\theta x, y/\theta)$ where $f(\cdot,\cdot)$ is any suitably regular bivariate density on the first quadrant. Then with the same transformation the joint density of (u,v) is

$$(B.2) \qquad g(u,v;\tau) = e^{v-u}f(e^{\tau-u},e^{v-\tau}) = e^{z}e^{-y}f(e^{-y},e^{z}),\ y = u-\tau,\ z = v-\tau$$

with the previously mentioned ancillary, which would not in general be reducible.

A second generalization, mentioned by Fisher (1973), p. 175, (but omitted in Section 8 above) takes $f(x,y;\theta) = \theta\phi e^{-\theta x-\phi y}$ with $\phi = \theta^{s}$. This reduces to a location model under the transformation $\tau = \log\theta$, $u = -\log x$ and $v = -(1/s)\log y$. Evidently this model itself generalized as above to $\theta^{s+1}f(\theta x,\theta^{s}y)$.

Appendix C

Inference With Known Coefficient of Variation

Hinkley (1977) has considered inference about μ when the parent population is $N(\mu, c^2\mu^2)$ where the coefficient of variation c is known. If we assume with Hinkley that $\mu > 0$ then the density and CDF are

$$(C.1) \qquad f(x;\mu) = (1/c\mu)\phi[(x-\mu)/c\mu]$$

and

$$(C.2) \qquad F(x;\mu) = \Phi[(x-\mu)/c\mu] = G(x/\mu)$$

where ϕ and Φ are the standard normal density and CDF and where

$$(C.3) \qquad G(\lambda) = \Phi(\frac{\lambda}{c} - \frac{1}{c}).$$

Let us consider a generalization in which ϕ, Φ, G are replaced by ψ, Ψ, H, where ψ is an arbitrary density with support $(-\infty,\infty)$. If x has density (1) with ψ substituted for ϕ then

$$P(x \leq 0) = \Psi(-1/c) = H(0) = q, \text{ say.}$$

Thus the indicator function $I(x)$ which equals 1 for $x \leq 0$, 0 for $x > 0$ is an ancillary statistic. For n observations the corresponding indicators $I_1,\ldots,I_n$ are jointly ancillary.

Returning to one observation, given that $x > 0$ the CDF of x is

$$F_+(x;\mu) = \frac{1}{p}(F(x;\mu) - q)$$

where $p = 1 - q$, from which we get the induced density

(C.5) $$g(\mu|x) = -\frac{\partial}{\partial\mu}F(x;\mu) = \frac{x}{pc\mu^2}\psi(\frac{x}{c\mu} - \frac{1}{c}) = \frac{x}{p\mu}f(x;\mu) .$$

Similarly, given that $x < 0$, the induced density is

(C.6) $$g(\mu|x) = -\frac{x}{p\mu}f(x;\mu).$$

The last two expressions agree with a posterior corresponding to the prior $d\mu/\mu$.

Next consider two observations, x_1, x_2. If both are positive we have two independent observations from F_+, that is, two observations from a scale family, and it is known from the Fisher-Pitman theory that the fiducial distribution obtained by conditioning on the ancillary x_1/x_2 equals the posterior for prior $d\mu/\mu$. A similar argument applies if both are negative. If $x_1 > 0$ and $x_2 < 0$ then x_1/x_2 is again ancillary and we again get a posterior corresponding to the same prior.

Finally suppose that of n observations, r are negative and $s = n - r$ are positive. Without loss of generality we may suppose the first r are negative. If $y_i = \log|x_i|$, $\tau = \log\mu$, the joint density of the y's has the form

(C.7) $$\prod_{i=1}^{r} h_1(y_i - \tau) \prod_{i=r+1}^{n} h_2(y_i - \tau)$$

a generalized location model which falls within the generalized Fisher-Pitman theory described in Appendix A. The spacings $(y_1-y_2, y_2-y_3, \ldots, y_r-y_{r+1}, \ldots, y_{n-1}-y_n)$ are ancillary, and the induced distribution of τ is a posterior corresponding to a uniform prior, the separate cases need not be distinguished in stating that the induced distribution of μ in simply the posterior distribution corresponding to prior $d\mu/\mu$.

Appendix D

Sprott's Ancillary

In Sprott's (1961) example, $x_1 \sim N(n\theta,n)$ and $x_2 \sim G(m,ce^{k\theta})$. Thus $x_1/n \sim N(\theta,1)$ and $y_2 \equiv ce^{k\theta}x_2 \sim G(m,1)$. We get $\log y_2 = \log c + k(\theta + (1/k)\log x_2)$. Since the distribution of y_2 is free of θ we see that θ is a location parameter for $z_2 \equiv (1/k)\log x_2$. But θ is also a location parameter for $z_1 \equiv x_1/n$. Therefore by the location parameter theory of Appendix A, an ancillary statistic is $z_1 - z_2 = x_1/n-(1/k)\log x_2$ (as Sprott showed by a different argument).

Appendix E

The Lindley Distribution

The Lindley distribution was originally presented (Lindley, 1958) as an example satisfying Condition B (B-regularity; see D1 in Section 2) but not Condition A.

We write $x \sim \text{Lind}(\theta)$ if x has density

$$\text{(E.1)} \qquad f(x;\theta) = \theta^2(\theta + 1)^{-1}(x + 1)e^{-\theta x} \qquad x > 0,\ \theta > 0.$$

The CDF is

$$\text{(E.2)} \qquad F(x,\theta) = 1 - e^{-\theta x}[1 + \theta x/(\theta + 1)].$$

Given one observation x, the MLE $\hat{\theta}$ is the value of θ satisfying

$$\text{(E.3)} \qquad x = \psi(\theta) = \frac{2}{\theta} - \frac{1}{\theta+1} = \frac{\theta+2}{\theta(\theta+1)}.$$

Thus the MLE has CDF

$$\text{(E.4)} \qquad \begin{aligned} P\{\hat{\theta} \leq w\} &= P\{\psi(\hat{\theta}) \geq \psi(w)\} \\ &= P\{x \geq \psi(w)\} \\ &= 1 - P\{x \leq \psi(w)\} \\ &= \{\exp(-\theta\psi(w))\}\{1 + \theta\psi(w)/(\theta + 1)\}. \end{aligned}$$

With θ as abscissa and w as ordinate, vertical sections (θ fixed) of this last function give values of the CDF of θ, horizontal sections (w fixed) give one minus the induced CDF of θ, and thus conditional confidence limits for model EI2 of Section 8 when $u = 1$.

Appendix F

A Conditional Exponential Model

Barndorff-Nielsen (1980) calls attention to the following example of an exponential family of densities which appeared in Lloyd and Saleem (1979):

$$\text{(F.1)} \qquad f(x,y;\alpha,\theta) = \frac{(\alpha\theta-1)^n}{\Gamma(n)} I_{n-1}(2\sqrt{xy})(xy)^{(n-1)/2} e^{-\alpha x-\theta y} \quad x > 0,\ y > 0.$$

Here $\alpha > 0$, $\theta > 0$, $\alpha\theta > 1$, and $I_{n-1}(\cdot)$ is a Bessel function. The marginal distributions are both gamma, and so (F.1) can be called a correlated bivariate gamma distribution. For our purposes it will suffice to consider the special case $n = 1$, which gives

$$\text{(F.2)} \qquad f(x,y;\alpha,\theta) = (\alpha\theta-1)I_o(2\sqrt{xy})e^{-\alpha x-\theta y} \quad x > 0, \quad y > 0.$$

In this case the Bessel function can be represented by

$$\text{(F.3)} \qquad I_o(2\sqrt{u}) = \sum_{j=0}^{\infty} u^j/(j!)^2.$$

Termwise integration with respect to y gives the marginal distribution $x \sim \text{Exp}(\alpha - \theta^{-1})$. From this, the conditional density of y given x is

$$\text{(F.4)} \qquad f(y|x;\theta) = \theta I_o(2\sqrt{xy})e^{-(x/\theta)-\theta y}.$$

The last expression is free of α (as was to be expected from Lehmann-Scheffe theory of exponential families) and gives us the basis for tests and confidence intervals for θ (when α is a nuisance parameter) having known optimal properties.

To view the problem as a one-parameter curved exponential family in the same sense of Efron (1975), we can fix some function of α,θ. In particular, take any $c > 0$, and suppose α,θ are restricted to the hyperbolic contour $\alpha - \theta^{-1} = c$. Then x,y have a bivariate density

$$\text{(F.5)} \qquad f(x,y;\theta) = c\theta I_o(2\sqrt{xy})e^{-cx-(x/\theta)-\theta y}$$

$$= c\theta e^{-cx-(x/\theta)-\theta y} \sum_{j=0}^{\infty} x^j y^j/(j!)^2 \qquad x > 0, \quad y > 0$$

Here the parameter θ is retained as a convenient coordinate measuring the location on the hyperbolic contour determined by the known value c. The construction has been arranged so that x is ancillary: $x \sim \text{Exp}(c)$. From either (F.4) or (F.5) we get the log likelihood

$$\text{(F.6)} \qquad \log f = \text{constant} + \log\theta - (x/\theta) - \theta y$$

and the score function

$$\text{(F.7)} \qquad \partial \log f/\partial\theta = (1/\theta) + (x/\theta^2) - y.$$

From (F.7) we can easily obtain the conditional and unconditional Fisher information:

$$\text{(F.8)} \qquad I_{y|x}(\theta) = \theta^{-2} + 2x\theta^{-3}, \quad I_{x,y}(\theta) = \theta^{-2} + 2c^{-1}\theta^{-3}.$$

Since the conditional information does not factor into a function of θ times a function of x, we conclude from Proposition 2, Section 6, that condition P2T does not hold.

Barndorff-Nielsen (1979b) has pointed out a curious property of the score function (F.7) (and others arising from exponential families): The value of y does not change the shape of the function, but only gives a vertical translation. (Horizontal translations are more familiar.)

For any fixed c the conditionality principle prescribes the use of the conditional distribution (F.4), and the confidence intervals we get in this way are the same for any c and correspond to the Lehmann-Scheffe solution. The Lehmann-Scheffe theory applies to exponential families generally. On the other hand the development via the curved exponential family requires the distribution of x to be parameter-free along a curve in the parameter space. The secret of success here is that the marginal distribution of x depends only on a single parameteric function, namely $Ex = (\alpha - \theta^{-1})^{-1}$. This corresponds to x being a "cut" in the sense of Barndorff-Nielsen (1979a), p. 50. For the general distribution (F.1), x is a cut because for any $\alpha_1, \theta_1, \theta_2$ (α_2 does not enter) we can find α, θ such that

(F.9) $$f_1(x;\alpha_1,\theta_1)f_2(y|x;\theta_2) = f(x,y;\alpha,\theta).$$

The conditional distribution (F.4) superficially appears to depend on both θ and x, but these collapse to a single value. From (F.4) we find that the conditional distribution of y given x can be expressed as

(F.10) $$\{y|x\} \sim (1/2\theta)\chi_2^2(x/\theta)$$

where χ_2^2 is the noncentral chi square distribution defined in Section 7.

From (F.7) we get the maximum likelihood estimator

$$\text{(F.11)} \qquad \hat{\theta} = [1 + (1 + 4xy)^{\frac{1}{2}}]/(2y)$$

which is an increasing function of x for fixed y and a decreasing function of y for fixed x.

Since the MLE $\hat{\theta}$ is monotone in y for fixed x, the same induced distribution of θ results from use of y or of $\hat{\theta}$, conditional on x. If $z_{\lambda,\gamma}$ denotes the γ percentile of $z_\lambda \sim \chi_2^2(\lambda)/2$, then from (F.10) the γ percentile of the distribution of y given x, is $(1/\theta)z_{x/\theta,\gamma}$. From this we get confidence limits for θ, that is, the induced distribution of θ, by the usual inversion: $\bar{\theta}_\gamma(x,y)$ is the value of θ for which $z_{x/\theta,\gamma} = \theta y$. The implicit solution for θ gives the upper confidence limit as a function of (x,y).

For our purposes it is preferable to work with $(x,\hat{\theta},\theta)$ rather than (x,y,θ), in order to consider the contours of the conditional CDF $F(\hat{\theta}|x;\theta)$ in the $(\theta,\hat{\theta})$ plane. Since $\hat{\theta}$ is monotone decreasing in y, put $\gamma' = 1 - \gamma$ and use (F.11) to get

$$\text{(F.12)} \qquad \bar{\theta}_{\gamma'} = \{1 + \sqrt{1 + 4xy_\gamma}\}/(2y_\gamma)$$

where y_γ is the γ percentile of the distribution of y given x,θ. Substituting $y_\gamma = (1/\theta)z_{x/\theta,\gamma}$ gives

$$\text{(F.13)} \qquad \bar{\theta}_{\gamma'} = x\{1 + \sqrt{1 + 4\lambda z_{\lambda,\gamma}}\}/(2\lambda z_{\lambda,\gamma})$$

where $\lambda = x/\theta$. This formula allows us to calculate the percentiles of the distribution of $\hat{\theta}$ given x and θ from the percentiles of the noncentral chi-squared variate $2z_\lambda$.

We were unable to find tabled values of noncentral chi-square to suit our needs. Therefore values for $\lambda = 1/16, 1/8, \ldots, 4, 8$ and $\gamma = 0.1(0.2)0.9$ were calculated on an Ohio Scientific C4P home computer. A few sample values are given in Table 1. From these we obtained $\bar{\theta}_\gamma$, percentiles of the conditional distribution of $\hat{\theta}$ given θ and x. Representative values are given in Table 2.

Figure 1 is a log-log plot of the 0.1 and 0.9 contours of $F(\hat{\theta}|x,\theta)$. If these were 45 degree lines, θ would be a scale parameter. In fact the departure from 45 degree lines is rather small so that θ is approximately a scale parameter. One $\gamma = 0.7$ contour is shown to demonstrate that it crosses the 0.9 contours. Accordingly x is not an "exact index of precision."

Figure 1

Below the diagonal line ($\hat{\theta} = \theta$) are the contours $F(\hat{\theta}|x;\theta) = 0.1$ for $x = 1/16, 1/8, \ldots, 8$. The curves above the diagonal are contours $F(\hat{\theta}|x;\theta) = 0.9$ for the same x values. Confidence intervals become shorter as x increases. The dashed line is the contour $F(\hat{\theta}|1/2,\theta) = 0.7$.

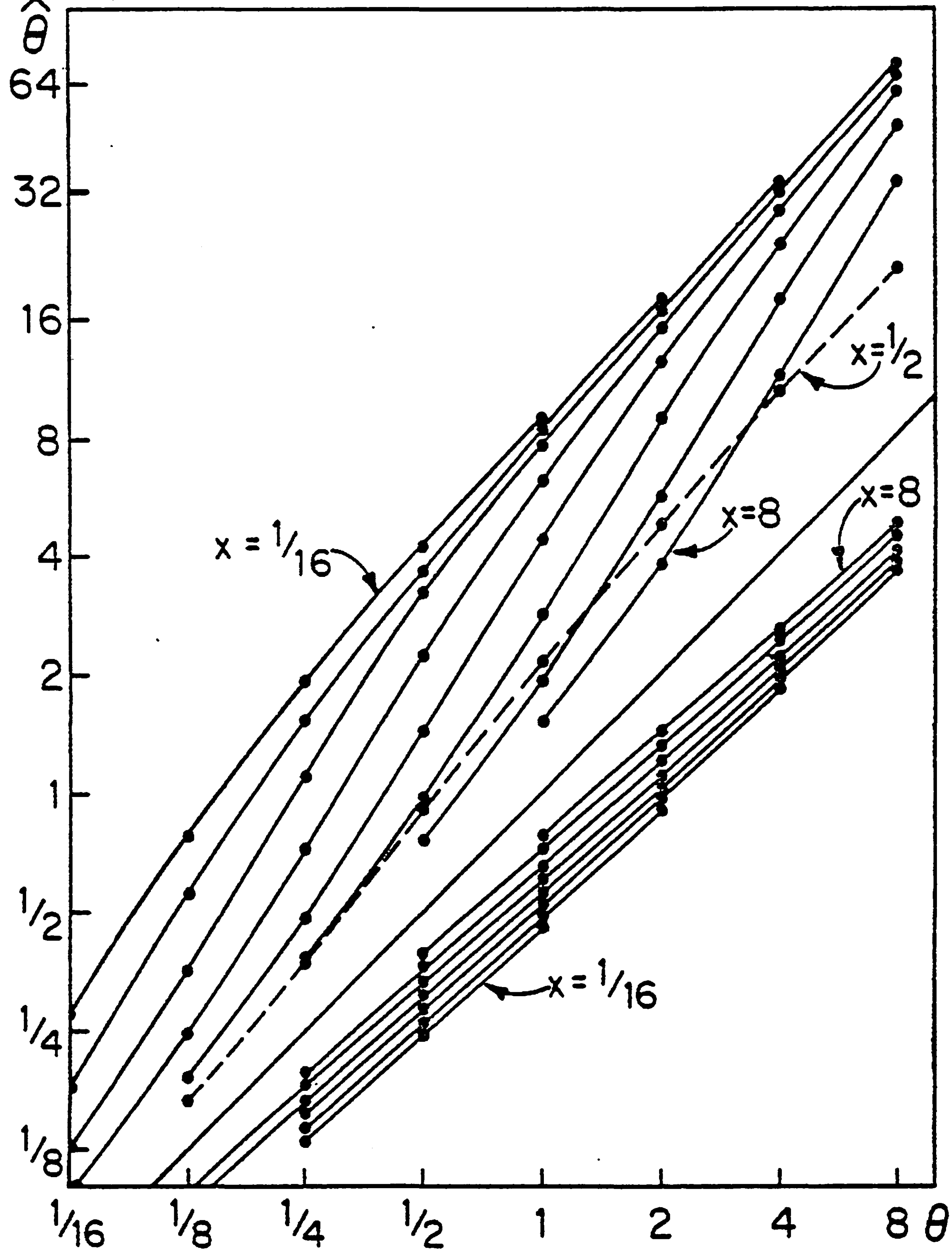

Table 1

Values of $z_{\lambda,\gamma}$ where $z \sim \chi_2^2(\lambda)/2$

λ	$\gamma = 0.1$	$\gamma = 0.9$
0.0625	.112	2.45
0.125	.119	2.59
0.25	.135	2.86
0.5	.172	3.39
1	.274	4.34
2	.595	6.15
4	1.577	9.07
8	4.110	14.58

Table 2

Percentiles of conditional distribution of $\hat{\theta}$

	$\gamma = 0.1$				$\gamma = 0.9$			
	$\theta = 1/16$	1/4	1	4	$\theta = 1/16$	1/4	1	4
$x = 1/16$	.038	.130	.464		.280	1.91	9.0	
1/4	.045	.152	.518	1.86	.121	1.12	7.65	35.9
1		.180	.609	2.07		.485	4.47	30.6
4			.722	2.44			1.94	17.9

REFERENCES

[1] Barnard, G. A., and Sprott, D. A. (1971). A note on Basu's examples of anomalous ancillary statistics. Foundations of Statistical Inference (Godambe and Sprott, eds.), 163-176.

[2] Barndorff-Nielsen, Ole (1979a). Information and exponential families in statistical theory. New York: Wiley.

[3] Barndorff-Nielsen, Ole (1979b). Untitled unpublished manuscript.

[4] Barndorff-Nielsen, Ole (1980). Conditionality resolutions. Biometrika, 67, 293-310.

[5] Basu, D. (1959). The family of ancillary statistics. Sankhya, 21, 247-256.

[6] Basu, D. (1964). Recovery of ancillary information. Sankhya Series A, 26, 3-16.

[7] Buehler, Robert J. (1959). Some validity criteria for statistical inferences. Ann. Math. Statist., 30, 845-863.

[8] Cox, D. R. (1958). Some problems connected with statistical inference. Ann. Math. Statist., 29, 357-372.

[9] Cox, D. R. (1971). The choice between ancillary statistics. J. Roy. Statist., Soc. Ser. B, 33, 251-255.

[10] Cox, D. R. (1980). Local ancillarity. Biometrika, 67, 279-286.

[11] Cox, D. R. and Hinkley, D. V. (1974). Theoretical Statistics. London: Chapman and Hall.

[12] Efron, Bradley (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). Ann. Statist., 3, 1189-1242.

[13] Efron, Bradley and Hinkley, David V. (1978). Assessing the accuracy of the maximum likelihood estimates: Observed versus expected Fisher information. Biometrika, 65, 457-487.

[14] Fisher, R. A. (1925). Theory of statistical estimation. Proc. Cambridge Philos. Soc., 22, 700-725.

[15] Fisher, Sir Ronald A. (1973). Statistical methods and scientific inference (third edition). New York: Hafner.

[16] Fraser, D. A. S. (1961a). On fiducial inference. Ann. Math. Statist., 32. 661-676.

[17] Fraser, D. A. S. (1961b). The fiducial method and invariance. Biometrika, 48, 261-280.

[18] Godambe, V. P. and Sprott, D. A. (1971). Foundations of Statistical Inference. Toronto: Holt Reinhart and Winston.

[19] Graybill, Franklin A. (1976). Theory and Application of the Linear Model. Belmont, California: Duxbury Press.

[20] Hinkley, D. V. (1977). Conditional inference about a normal mean with known coefficient of variation. Biometrika, 64, 105-108.

[21] Hinkley, D. V. (1980). Likelihood as an approximate pivotal distribution. Biometrika, 67, 287-292.

[22] Hora, R. B. and Buehler, R. J. (1966). Fiducial theory and invariant estimation. Ann. Math. Statist., 37, 643-656.

[23] Kendall, M. G. and Buckland, W. R. (1957). A Dictionary of Statistical Terms. Edinburgh: Oliver and Boyd.

[24] Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. J. Roy. Statist., Ser. B, 20, 102-107.

[25] Lloyd, E. H., and Saleem, S. D. (1979). A note on seasonal Markov chains with gamma or gamma-like distributions. _J. Appl. Prob._, _16_, 117-128.

[26] Sprott, D. A. (1961). An example of an ancillary statistic and the combination of two samples by Bayes' theorem. _Ann. Math. Statist._, _32_, 616-618.